JURNAL EMACS
[Engineering, MAthematics and Computer Science] Vol.2 No.3 September 2020: 133-139 e-ISSN: 2686-2573

Semantic Segmentation for Aerial Images: A Literature Review

Yongki Christian Sanjaya[1], Alexander A S Gunawan[2], Edy Irwansyah[3]

[1,2,3] Computer Science Department, School of Computer Science,
Bina Nusantara University,
Jakarta, Indonesia 11480
yongki.sanjaya@binus.ac.id; aagung@binus.edu; eirwansyah@binus.edu

Abstract - *Semantic image segmentation is one of the fundamental applications of computer vision which can also be called pixel-level classification. Semantic image segmentation is the process of understanding the role of each pixel in an image. Over time, the model for completing Semantic Image Segmentation has developed very rapidly. Due to this rapid growth, many models related to Semantic Image Segmentation have been produced and have also been used or applied in many domains such as medical areas and intelligent transportation. Therefore, our motivation in making this paper is to contribute to the world of research by conducting a review of Semantic Image Segmentation which aims to provide a big picture related to the latest developments related to Semantic Image Segmentation. In addition, we also provide the results of performance measurements on each of the Semantic Image Segmentation methods that we discussed using the Intersectionover-Union (IoU) method. After that, we provide a comparison for each semantic image segmentation model that we discuss using the results of the IoU and then provide conclusions related to a model that has good performance. We hope this review paper can facilitate researchers in understanding the development of Semantic Image Segmentation in a shorter time, simplify understanding of the latest advancements in Semantic Image Segmentation, and can also be used as a reference for developing new Semantic Image Segmentation models in the future.*

Keywords: *Semantic Image Segmentation; Computer Vision*

I. INTRODUCTION

Semantic Image Segmentation has an important role in Computer Vision problems. Semantic image segmentation is the process of understanding the role of each pixel in an image. Since Fully Convolutional Networks (FCN) [1] which popularized the Convolutional Neural Networks (CNN) architecture in predicting densities without fully connected layers were introduced, semantic image segmentation has become famous.

Over time, the rapid growth of the technological world has produced various architectural models that have emerged to solve the Semantic Image Segmentation problem. In addition, semantic image segmentation has also been used or applied in many domains such as medical areas and intelligent transportation [2]. In medical areas, semantic image segmentation is used to detect brains and tumors [3], and detect and track medical instruments in operations [4]. Whereas in intelligent transportation, semantic image segmentation is used to detect road signs [5], colon crypts segmentation [6], land use and land cover classification [7].

With this rapid development, a broad review of Semantic Image Segmentation is very important for developing new ideas in future research. Our motivation in making this paper is to contribute to the world of research by conducting a review of Semantic Image Segmentation which aims to provide a big picture related to the latest developments related to Semantic Image Segmentation. In addition, we also provide the results of performance measurements on each of the Semantic Image Segmentation methods that we discussed using the Intersectionover-Union (IoU) method [8]. After that, we provide a comparison for each semantic image segmentation model that we discuss using the results of the IoU and then provide conclusions related to a model that has good performance. With the presence of this paper, we hope this will provide convenience for researchers in this field so that the new Semantic Image Segmentation architectural model can be developed.

We use the Traditional Review method in this study. Semantic Image Segmentation is a problem that requires the right method to solve it. Therefore, there are relatively few research papers that can be used as references. Thus, the use of traditional review methods is the right choice. This paper is organized as follows: It begins by giving a summary of models used in this study as to solve the problem of image

segmentations in Section II. A summary of quality measures and datasets which are used here in Section III. A summary of evaluation results of each models as well as the results of discussion follows in Section IV, as well as the conclusion in Section V.

II. METHODS

In this paper, we provide a big picture of several models that can solve the problem of semantic image segmentation. Some of these models include:

2.1. Object-Contextual Representations for Semantic Segmentation

This model proposes the use of Object-Contextual Representations (OCR) + HRNetV2 [9] methods to solve semantic image segmentation problems. The OCR method provides a simple but effective approach, characterizing pixels by exploiting the appropriate object class representation. The focus in discussing the Semantic Image Segmentation problem in this method is the context aggregation strategy where the motivation is the class label assigned to one pixel is the object category that the pixel belongs to.

The first process carried out by this method is to study the area of objects by dividing contextual pixels into a set of soft object regions with each corresponding to the class under the supervision of the ground-truth segmentation. Second, estimating the representation of the object's area by combining pixel representations located in the object's region. Third, calculate the relationship between each pixel and each object region, and add representation of each pixel with the object-contextual representation which is a weighted aggregation of all object area representations according to their relationship to pixels.

There are two things that make this method unique to solving Semantic Image Segmentation problems. First, in terms of the conventional multi-scale context schemes, OCR distinguishes the same-object-class contextual pixels from the different object-class contextual pixels. Second, from the relational context schemes, OCR arranges contextual pixels into object regions and exploits the relationship between pixels and object regions.

This method has a good performance in terms of solving Semantic Image Segmentation problems. The OCR approach outperforms other approaches, such as DANet [10]. After being evaluated using various benchmarks namely Cityscapes (83,7% mIoU), ADE20K (45,66% mIoU), LIP (56,65% mIoU), Pascal Context (56,2% mIoU), and COCO-Stuff (40,5% mIoU), good and competitive performance results were obtained.

2.2. Understanding Convolution for Semantic Segmentation

This model proposes the use of the DUC-HDC [11] method to solve the semantic image segmentation problem. The DUC-HDC method is a development of deep convolutional neural networks (CNNs) which previously contributed well to semantic segmentation systems. DUC-HDC is a method that can improve pixel-based semantic segmentation by manipulating convolution related operations that have theoretical and practical values.

The uniqueness of this method lies in the decoding and encoding section. On the decoding side, this method proposes dense upsampling convolution (DUC) to get good accuracy at the pixel level and capture and decode more detailed information that will generally be lost in bilinear upsampling. On the other hand, in encoding this method proposes a simple hybrid dilation convolution (HDC) framework. HDC has several two advantages. First, this framework effectively enhances network receptive fields (RF) to collect global information. Second, this framework can alleviate gridding problems caused by the standard dilated convolution operation.

The performance of DUC-HDC in Semantic Image Segmentation problem solving is not as good as the performance of the OCR [9] method. However, it has competitive performance from the OCR method. This is evidenced by the evaluation conducted using the Cityscapes dataset, the resulting performance was 80.1% mIoU.

2.3. Pyramid Scene Parsing Network

The Pyramid Scene Parsing Network (PSPNet) [12] model provides a special approach that is quite stable, solve representative failure cases by applying the Fully Connected Network method that is useful in every decomposition event. The focus in discussing the Semantic Image Segmentation problem in this method is pyramid collection as a module for a more effective global context where the motivation is to expand pixel-level features into a single global pyramid pooling designed specifically for pixel prediction.

Each process carried out in this method is to review the latest progress in the scene Segmentation Parsing and Semantic Image Segmentation which functioned in pixel-level prediction to replace the fully connected layer in the classification of Layer Convolution so that it can Enlarge the receptive field of Neural Networks by using Dilated Convolution to propose a rough to smooth structure with a deconvolution network in learning the Segmentation Mask then Combining the multi-scale features of the Fully Connected Network and the Dilated Network to cover higher layers which contain more Semantics and fewer locations which increases two-way performance.

In solving the Semantic Image Segmentation problem there are three processes that differentiate and really help FCN, the first process is the existence of a pyramid scene to embed difficult context features into the pixel prediction framework, then in optimizing an effective strategy for ResNet based on pixel losses and then build parsing and semantic segmentation where all implementations can be included for practical systems in pixel prediction.

In the FCN Method performance is good enough so that to compare the other approaches is quite maximal, such as the ShelfNet [13] approach, in the discussion of the

Dataset that has been carried out, namely ADE20K (44.94% mIoU) and Cityscapes (80.2% mIoU), produced a very good evaluation.

2.4. Improving Semantic Segmentation via Video Propagation and Label Relaxation

This model proposes the use of the DeepLabV3Plus + SDCNetAug [14] method to solve the semantic image segmentation problem. The Video-Prediction Method takes a model approach that focuses on accuracy, characterize samples into training sets in video predictions to improve network accuracy in discussing Semantic Image Segmentation problems, in this method is the strategy of taking the ability of the video prediction model where the motivation is to predict future frames and future labels.

The first process carried out by this method is to create a new training sample to be propagated label with the original future frame called the Propagation Label. Second, create a new training sample to be propagated label with corresponding propagated images that use each other's past labels and frames that work together in prediction models, the resulting image-label pairing will increase alignment higher called Joint image-label Propagation.

There are three things that make this method unique to solving Semantic Image Segmentation problems. First, in terms of Patch Matching which tends to be sensitive to patch size and threshold values, Patch Making distinguishes class statistics with a variety of knowledge and then ranks according to the sensitivity of the patch. Second, there is Optical Flow which tends to be very accurate in accuracy, Optical Flow reduces miss-alignments propagated labels with corresponding frames. Third, in terms of Boundary Handling, combining constraints as control pixel boundaries in video predictions.

The used network architecture is based on DeeplabV3Plus[15] which use encoder-decoder networks like U-Net combined with atrous convolution. It attempts to take advantages of both methods which to faster computation with the encoder-decoder networks while applying atrous convolution to extract denser feature maps. U-Net is an example of encoder-decoder network. It has a symmetrical U-shaped network which gained its name from. The left side of network consists on feature extraction layer while the right side is for upsampling with bottleneck layers in the middle side. Atrous convolution on the other hand is a powerful tool to explicitly control the resolution of features computed by deep convolutional neural networks. It is a standard convolution with added stride rate which allowed the network to enlarge the filter's field-of-view.

This method has good performance and significant accuracy in terms of solving Sematic Image Segmentation problems. The DeepLabV3Plus + SDCNetAug approach outperforms other approaches, for example InPlaceABN [16]. This is evidenced by the evaluation conducted using the Cityscapes dataset, the resulting performance was 83.5% mIoU. However, this result is still slightly lower compared to OCR [9].

2.5. Context Prior for Scene Segmentation

The Context Prior Network [17] method conducts affinity monitoring for context prior which is useful for building an ideal affinity in the form of image and corresponding ground truth. Focus on Semantic Image Segmentation which provides a good strategy and stable accuracy, so this method builds context prior layer to capture the intraclass and interclass contextual dependencies explicitly, then context prior embedded in the context prior layer with an explicit affinity loss to supervise the learning process.

In carrying out the process, this method explains two paths for capturing contextual dependencies in which this Context Aggregation studies the capture of undesirable contextual dependencies without explicitly distinguishing the difference of different contextual relationships and then Attention Mechanism studies the leading to an undesirable context aggregation.

As for what makes this approach more supportive to solve the Semantic Image Segmentation problem in the form of an effective context design Priority network for scene segmentation, which contains a backbone network and a context prior layer.

Therefore, this method has a good performance in terms of solving Semantic Image Segmentation problems compared to the others as the CPN approach outperformed the PSPNet [12] approach. The results of evaluation of the use of various datasets, namely ADE20K (46.3% mIoU) and Cityscapes (81.3% mIoU), obtained significant results.

2.6. ShelfNet for Fast Semantic Segmentation

In the use of the Shelfnet method [13], which is useful for image segmentation semantics, it has the latest artificial form that is fast and has good enough accuracy so that Shelfnet has a number of pairs of connection encoder-decoder connections to pass through each spatial level, which looks like a rack with multiple columns. The essence of semantic image segmentation provides good accuracy and information so that the use of the Shelfnet Structure can be seen as multiple ensembles both inside and outside the path, which can increase accuracy.

When running the process, the method used at the same time can reduce the computational burden by reducing the channel number in the use of segmentation racks that have the weight of two convolutional sections in 1 residue block, functioning to reduce the number of parameters without losing accuracy.

As for what makes this approach more supportive for solving Semantic Image Segmentation problems in the form of feature maps encoded by various stages of the backbone that are inserted into the segmentation rack then the more paths in the feature map the more information that can be used in the Shelfnet encoder-decoder . Compared to the BiSeNet method [18], ShelfNet can have a speed of inference 4 × faster with the same accuracy. Shelfnet can activate applications in tasks that demand speed such as understanding street scenes for autonomous driving.

Therefore, this method has a good performance in terms of solving the Semantic Image Segmentation problem compared to the others because the Shelftnet approach outperforms the BiSeNet approach. Evaluation results of various data set uses, namely, Cityscapes (79.0% mIoU) and Pascal Context (48.4% mIoU).

2.7. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs

In the use of the DeepLab-CRF method (Resnet-101) [19] which is useful for semantic image segmentation is done practically and significantly. In carrying out a process of atrous spatial pyramid pooling (ASPP) that is useful for dynamically segmented objects at various scales, there are convolutional features that enter layers that filter several levels of sampling effectively, so they can capture objects and images at several scales. Then, increasing object boundary localization by combining a combination of max-pooling and downsampling using a probabilistic Graphic model and the DCNN method. The combination of max-pooling and DeepLab-CRF downsampling (Resnet-101) to achieve variable data types with accurate location.

The focus of semantic segmentation in the use of Deeplab-CRF (Resnet-101) in the form of bottom-up image segmentation with various classifications, the use of convolution features in labeling solid images can be paired with segmentation independently so that directly when pixels at a solid level can eliminate segmentation that is solid not reach the same level. In this approach, it will be useful to support solutions to semantic segmentation image problems better by observing convolution upampled filters as a powerful tool for predicting things accurately, and upampled filters can also control the inference of image content in real layer features.

Therefore, this method has a good performance in solving semantic image segmentation problems compared to the others because the DeepLab-CRF (Resnet-101) approach is superior to the FCRN method [20]. The evaluation results in the use of data sets namely, PASCAL-Context which reached 45.7% mIoU and Cityscapes which reached 70, 4% mIoU.

III. RESULT AND DISCUSSION

3.1. Evaluation

The performance measurement results for each of the Semantic Image Segmentation methods that we discuss using the Intersectionover-Union (IoU) method [8]. The Intersectionover-Union (IoU) method [8] is a standard performance measure commonly used for semantic image segmentation problems. The IoU size gives a similarity between the predicted region and the ground-truth region for the object in the image and is defined as the size of the intersection divided by the union of the two regions. The IoU measure can take into account the problem of class imbalance that is usually present in setting such problems.

$$IoU = \frac{TP}{FP + TP + FN}.$$

Figure. 1 Formula IoU, where TP (True Positive), FP (False Positive), FN (False Negative)

From fig. 1, we see that IoU is a measure based on count, whereas, the output of each semantic image segmentation model is a probability value that represents the likelihood of pixels being part of the object. Therefore, we cannot accurately measure IoU directly from the network output. We need to adjust IoU measurements using ability values.

At the end, we provide a comparison related to each semantic image segmentation model that we discussed, then give a conclusion about a model that has good performance in solving the semantic image segmentation problem.

3.2. Discussion

The discussion we conducted related to the model discussed earlier resulted in some data including:

a. Advantages and Disadvantages of the Method in Semantic Image Segmentation

Table 1. Advantages and Disadvantages of the Method in Semantic Image Segmentation

No	Method	Advantage	Disadvantage
1.	(OCR) + HRNetV2	OCR achieved the best performance of 83.7% mIoU for the Cityscapes dataset	In the ADE20K test, the performance was still less than APCNET
2.	DUC-HDC	DUC-HDC is effective for use in various semantic segmentation tasks	The results of ResNet-DUC-HDC-fine are smaller compared to ResNet-DUC-HDC-coarse in the cityscapes test set
3.	PSPNet	Pyramid tissue that is effective in complex understanding Additional contextual information provisioning features	Lack of decomposition of scenes in society. The technique used is still small
4.	DeepLab V3Plus + SDCNetAug	An effective video prediction-based data synthesis method for enhancing training tools for semantic segmentation Introducing a joint propagation strategy reduce miss-alignment in the synthesized sample Presents a new limit relaxation technique for reduce label noise	Allows expensive prices inside dataset collection It still lacks an increase in accuracy in the target Duty Less publicly available to the research community

5.	CPN	CPNet shows good performance for ADE20K datasets compared to other methods	In terms of Picture Accuracy, CPNet is still underperforming from SAC method
6.	DeepLab-CRF	That are experimentally shown to have substantial practical merit	With this strategy is that the training process is not ideal
7.	ShelfNet	Having several pairs of branch encoder-decoders by passing adjacent branches is useful to enable multiple paths of information flow and achieve high accuracy and can reduce the number of parameters without the need to lose accuracy	Still not stable enough to achieve high accuracy Reduce the number of parameters so that the speed is unstable

b. Comparison State-Of-The-Art Methodologies In Semantic Image Segmentation

In table 2, we list the performance of each semantic image segmentation method that we have discussed in this review paper. The performance that we show is based on the measurement method we have used and explained earlier, namely The Intersectionover-Union (IoU). This IoU score will be a standard measure of the performance of each method in semantic image segmentation which is then averaged to become a mean-IoU (mIoU). The score will be used as a comparison of each method in semantic image segmentation. That way, from table 2, we can conclude that the OCR + HRNetV2 method is the best approach compared to other methods in solving the problem of semantic image segmentation in two datasets namely Cityscapes and Pascal-Context. In addition, for the ADE20K dataset, the best performance in this discussion is owned by CPN.

Table 2. State-of-the-art Performance models on Cityscapes and ADE20K

Method	Cityscapes (% mIoU)	ADE20K (% mIoU)	Pascal-Context (% mIoU)
OCR + HRNetV2	**83,7%**	45.66%	**56.2%**
PSPNet	80.2%	44.94%	-
DUC-HDC	80.1%	-	-
DeepLabV3Plus + SDCNetAug	83.5%	-	-
CPN	81.3%	**46.3%**	-
DeepLab-CRF	70.4%	-	45.7%
ShelfNet	79.0%	-	48.4%

IV. CONCLUSION

In this paper, we review several models that can be used in solving semantic image segmentation problems to simplify and accelerate understanding of the latest advances related to semantic image segmentation. Based on the results obtained, we can conclude that the Object-Contextual Representations model that uses the OCR + HRNetV2 method as a whole is the most successful and stable method to date. Therefore, solving the problem of semantic image segmentation in the future might consider this method to improve its performance again. Besides that, other promising and competitive approaches in solving semantic image segmentation are PSPNet, CPN, and DeepLabV3Plus + SDCNetAug.

REFERENCES

J. Long, E. Shelhamer and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," 2015.

X. Liu, Z. Deng and Y. Yang, "Recent progress in semantic image segmentation," 2018.

N. Moon, E. Bullitt, K. v. Leemput and G. Gerig, "Automatic Brain and Tumor Segmentation," in In 2002 International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2002.

G.-Q. Wei, K. Arbter and G. Hirzinger, "Automatic tracking of laparoscopic instruments by color coding," in In 1997 International Conference on Medical Robotics and Computer-Assisted Surgery (MRCAS), 1997.

S. Maldonado-Bascon, S. Lafuente-Arroyo, P. Gil-Jimenez, H. Gomez-Moreno and F. Lopez-Ferreras, "Road-Sign Detection and Recognition Based on Support Vector Machines," IEEE Transactions on Intelligent Transportation Systems, pp. 264-278, 2007.

A. Cohen, E. Rivlin, I. Shimshoni and E. Sabo, "Memory based active contour algorithm using pixel-level classified images for colon crypt segmentation," Computerized Medical Imaging and Graphics, pp. 150-164, 2015.

C. Huang, L. S. Davis and J. R. G. Townshend, "An assessment of support vector machines for land cover classification," International Journal of Remote Sensing, pp. 725-749, 2002.

M. A. Rahman and Y. Wang, "Optimizing Intersection-Over-Union in Deep Neural Networks for Image Segmentation," in In 2016 International Symposium on Visual Computing (ISVC), 2016.

Y. Yuan, X. Chen and J. Wang, "Object-Contextual Representations for Semantic Segmentation," 2019.

J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang and H. Lu, "Dual Attention Network for Scene Segmentation," 2019.

P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang, X. Hou and G. Cottrell, "Understanding Convolution for Semantic Segmentation," 2018.

H. Zhao, J. Shi, X. Qi, X. Wang and J. Jia, "Pyramid Scene Parsing Network," 2017.

J. Zhuang, J. Yang, L. Gu and N. C. Dvornek, "ShelfNet for Fast Semantic Segmentation," 2019.

Y. Zhu, K. Sapra, F. A. Reda, K. J. Shih, S. Newsam, A. Tao and B. Catanzaro, "Improving Semantic Segmentation via Video Propagation and Label Relaxation," 2019.

L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," 2018.

S. R. Bulò, L. Porzi and P. Kontschieder, "In-Place Activated BatchNorm for Memory-Optimized Training of DNNs," 2018.

C. Yu, J. Wang, C. Gao, G. Yu, C. Shen and N. Sang, "Context Prior for Scene Segmentation," 2020.

C. Yu, J. Wang, C. Peng, C. Gao, G. Yu and N. Sang, "BiSeNet: Bilateral Segmentation Network for Real-time Semantic Segmentation," in In 2018 The European Conference on Computer Vision (ECCV), 2018.

L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," 2017.

Z. Wu, C. Shen and A. v. d. Hengel, "Bridging Category-level and Instance-level Semantic Image Segmentation," 2016.

M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth and B. Schiele, "The Cityscapes Dataset for Semantic Urban Scene Understanding," 2016.

B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso and A. Torralba, "Scene Parsing through ADE20K Dataset," in 2017 Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

K. Gong, X. Liang, D. Zhang, X. Shen and L. Lin, "Look into Person: Self-supervised Structure-sensitive Learning and A New Benchmark for Human Parsing," in 2017 Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

H. Caesar, J. Uijlings and V. Ferrari, "COCO-Stuff: Thing and Stuff Classes in Context," 2018.

Z. Zhong, Z. Q. Lin, R. Bidart, X. Hu, I. B. Daya, Z. Li, W.-S. Zheng, J. Li and A. Wong, "Squeeze-and-Attention Networks for Semantic Segmentation," 2020.

S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang and P. H. S. Torr, "Conditional Random Fields as Recurrent Neural Networks," 2016.

M. Yang, K. Yu, C. Zhang, Z. Li and K. Yang, "DenseASPP for Semantic Segmentation in Street Scenes," in 2018 Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

Z. Wei, Y. Sun, J. Wang, H. Lai and S. Liu, "Learning Adaptive Receptive Fields for Deep Image Parsing Network," in 2017 Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

W. Wang, R. Yu, Q. Huang and U. Neumann, "SGPN: Similarity Group Proposal Network for 3D Point Cloud Instance," 2019.

G. Wang, P. Luo, L. Lin and X. Wang, "Learning Object Interactions and Descriptions for Semantic Image Segmentation," in 2017 Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

T. Takikawa, D. Acuna, V. Jampani and S. Fidler, "Gated-SCNN: Gated Shape CNNs for Semantic Segmentation," 2019.

T. Pohlen, A. Hermans, M. Mathias and B. Leibe, "Full-Resolution Residual Networks for Semantic Segmentation in Street Scenes," in 2017 Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

S. Mohajerani and P. Saeedi, "Cloud-Net+: A Cloud Segmentation CNN for Landsat 8 Remote Sensing Imagery Optimized with Filtered Jaccard Loss Function," 2020.

P. Luo, G. Wang, L. Lin and X. Wang, "Deep Dual Learning for Semantic Image Segmentation," in 2017 International Conference on Computer Vision (ICCV), 2017.

P. Li, Y. Xu, Y. Wei and Y. Yang, "Self-Correction for Human Parsing," 2019.

S. Kong and C. Fowlkes, "Recurrent Scene Parsing with Perspective Understanding in the Loop," 2017.

D. Dai and L. V. Gool, "Dark Model Adaptation: Semantic Image Segmentation from Daytime to Nighttime," 2018.

S. Choi, J. T. Kim and J. Choo, "Cars Can't Fly up in the Sky: Improving Urban-Scene Segmentation via Height-driven Attention Networks," 2020.

S. R. Bulò, G. Neuhold and P. Kontschieder, "Loss Max-Pooling for Semantic Image Segmentation," 2017.

S. M. Azimi, C. Henry, L. Sommer, A. Schumann and E. Vig, "SkyScapes – Fine-Grained Semantic Understanding of Aerial Scenes," in 2019 International Conference on Computer Vision (ICCV), 2019.

Z. Wu, C. Shen and A. v. d. Hengel, “Bridging Category-level and Instance-level Semantic Image Segmentation,” 2016.

L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff and H. Adam, “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation,” 2018.